

An Extended α -Helix and Specific Amino Acid Residues Opposite the DNA-binding Surface of the cAMP Response Element Binding Protein Basic Domain Are Important for Human T Cell Lymphotropic Retrovirus Type I Tax Binding*

(Received for publication, April 4, 1998, and in revised form, July 20, 1998)

Yong Tang[‡], Feng Tie[§], Imre Boros, Robert Harrod, Mark Glover[¶], and Chou-Zen Giam^{||}

From the Department of Microbiology and Immunology, Uniformed Services University, Bethesda, Maryland 20814 and the [¶]Department of Biochemistry, University of Alberta, Edmonton, Alberta T6G 2H7, Canada

The human T cell lymphotropic retrovirus type I (HTLV-I) trans-activator, Tax, interacts specifically with the basic-domain/leucine-zipper (bZip) protein, cAMP response element binding protein (CREB), bound to the viral Tax-responsive element consisting of three imperfect 21-bp repeat, each with a cAMP response element core flanked by G/C-rich sequences. Here, the minimal CREB-bZip necessary for Tax binding is shown to be composed of amino acid residues 280–341. The Tax-CREB interaction involves an uninterrupted and extended α -helix in CREB that spans most of its basic domain to include amino acid residues localized to the NH₂ terminus of the DNA binding region. Mutational analyses indicate that three residues, Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹ unique to this region of the CREB/activating transcription factor-1 subfamily of bZip proteins, constitute the contact surface for Tax. Amino acid substitutions in these positions had little impact on CREB-bZip binding to DNA but abrogated its binding to Tax. Each of the contact residues for Tax are spaced approximately two helical turns apart on the side of the bZip helix directly opposite to that of the invariant DNA-binding residues. Molecular modeling reveals the Tax-contact residues to be near the minor groove of the G/C-rich DNA in the 21-bp repeat. They most likely position Tax for minor groove contact with the G/C-rich sequences.

The human T lymphotropic retrovirus type I (HTLV-I)¹ trans-activator, Tax, stimulates viral transcription via three imperfect 21-bp repeat DNA elements in the HTLV-I U3 region (1–6). Each of the viral 21-bp-repeats contains a cAMP response element (CRE) core flanked by 5' G-rich and 3' C-rich

sequences. The 21-bp repeats, in collaboration with the cellular transcription factors, cAMP response element binding protein (CREB), CREB/activating transcription factor 1 (ATF-1) heterodimer, and to a lesser extent, ATF-1 homodimer, form nucleoprotein complexes uniquely capable of recruiting Tax into ternary complexes that mediate trans-activation (7–12). Whereas CREB binds to the CRE irrespective of DNA sequence context, the stable assembly of Tax, CREB, and DNA into a ternary complex and Tax-mediated trans-activation require the CRE and the 5' G-rich and 3' C-rich flanking sequences present in the viral 21-bp repeats (11–18). *In vitro* selection also indicates that DNA sequences bound preferentially by the Tax/CREB complex contain the CRE motif flanked by exceedingly long stretches of 5' G-rich or 3' C-rich sequences and bear striking resemblance to the HTLV-I 21-bp repeats (15). These results demonstrate directly that HTLV-I Tax is evolved principally toward activating transcription from the viral cis regulatory element, the HTLV-I 21-bp repeats.

The molecular basis for the DNA sequence specificity of Tax trans-activation is not well understood. DNase I footprinting and methylation interference performed on the Tax-CREB-21-bp-repeat complex revealed no protein protection of the G-rich and C-rich sequences (15). Further, naturally occurring or genetically engineered single nucleotide substitutions in the G-rich and C-rich sequences do not affect their function (6). Interestingly, recent data indicate that Tax may be involved in contacting the minor groove of the G/C-rich sequences (19).

The specificity for the flanking sequences requires a specific interaction between Tax and the basic region of CREB (20). Tax interacts with CREB via the latter's basic domain and amino acid residues in its immediate vicinity (11, 12, 20–22). Like CREB, Tax functions as a dimer (22, 23). Analyses of Tax mutants indicate that the CREB-binding domain is located in the NH₂ terminus, while the subunit dimerization domain resides in the middle section of Tax (10, 23, 24).

CREB and ATFs are prototypic bZip proteins noted for the common basic domain-leucine zipper structure responsible for sequence specific DNA recognition and intersubunit protein-protein interaction, respectively (25). Many members of the bZip family, including the mammalian CREB, ATF, c-Jun/c-Fos, and yeast GCN4, bind either the 7-bp AP-1 motif (TGA(C/G)TCA) or the related CRE motif (TGACGTCA). Remarkably, a third group of the families whose members include NF-IL6 (also known as C/EBP- β) and C/EBP, bind a set of DNA elements (T(T/G)NNGNAA(T/G) and CCAAT box, respectively) whose sequences differ significantly from the AP-1 site and CRE.

CREB becomes phosphorylated and activated as a function of the cAMP- or Ca²⁺-mediated signaling process (26–28). Both

* This work was supported by National Institutes of Health Grants RO1 CA48709 and RO CA/GM 75688. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

[‡] These two authors contributed equally to this work.

[§] Present address: Dept. of Genetics, Case Western Reserve University, School of Medicine, 10900 Euclid Ave., Cleveland, OH 44106.

[¶] To whom correspondence should be addressed: Dept. of Microbiology and Immunology, Uniformed Services University, 4301 Jones Bridge Rd., Bethesda, MD 20814. Tel.: 301-295-9624; Fax: 301-295-1545; E-mail: giam@bob.usuf2.usuhs.mil.

¹ The abbreviations used are: HTLV-I, human T cell lymphotropic retrovirus type I; ATF, activating transcription factor; bZip, basic domain-leucine zipper; CBP, CREB binding protein; C/EBP, CAAT/enhancer binding protein; CRE, cAMP response element; CREB, cAMP response element binding protein; GST, glutathione S-transferase; Tax, trans-activator from the pX region; EMSA, electrophoretic mobility shift assays; PCR, polymerase chain reaction; aa, amino acid; bp, base pair(s).

protein kinase A and Ca^{2+} /calmodulin-dependent kinases (I and IV) have been shown to phosphorylate CREB at Ser¹³³ to activate CRE/CREB-mediated transcription (26–28). A 265-kDa protein called CBP (CREB binding protein), and its cellular homologue, p300, specifically bind the Ser¹³³-phosphorylated form of CREB (29) and function as transcriptional co-activators of CREB (30). Kwok *et al.* (31) have shown recently that Tax interacts directly with the CBP and p300. These data have recently been confirmed and extended (32). Indeed, in *in vitro* protein/DNA binding reactions, the inclusion of CBP/p300 results in a highly stable quaternary nucleoprotein complex containing the 21-bp repeat/CREB/Tax/CBP-p300 (see “Results” and Ref. 32). In essence, Tax functions as a virus-specific link to connect the transcriptional co-activator, CBP/p300, and possibly other cellular transcription factors in a signal-independent manner to CREB/ATF-1 assembled on the viral 21-bp repeats. This allows HTLV-I viral gene expression to proceed in the absence of cellular activation.

X-ray structures of the bZip domain of GCN4 and the bZip heterodimer of c-Jun/c-Fos indicate that the dimerization of bZip proteins results from a coiled-coil structure composed of two intertwined α -helices, each derived from one subunit of the homo- or heterodimer (33–36). The leucine zipper and the basic domain form a continuous, uninterrupted α -helix. The basic domains extend from the leucine zipper like two prongs of a fork to engage DNA in the major groove (33–36). Many of the invariant amino acid residues (*boldface letters* in Fig. 1) in the basic domain are responsible for sequence-specific contacts with DNA (33–36). In this study, we show that an α -helix extending beyond the DNA-binding domain of CREB bZip is required for Tax binding. Along this extended helix, three amino acid (aa) residues, Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹, located on the opposing side of the invariant DNA recognition residues form the contact surface for Tax. The Tax-contact residues span a distance of approximately 23 Å along most of the length of the basic domain and beyond. Interestingly, these three amino acid residues point away from the major groove of the DNA and are posed adjacent to the minor groove of the G/C-rich sequences flanking the CRE motif, which we and others have shown to be critical for Tax binding (11, 12, 15, 19). Our data lend further support to the notion that Tax is involved in minor groove DNA contact with the G/C-rich flanking sequences after its binding to CREB.

EXPERIMENTAL PROCEDURES

Protein Expression and Purification—Expression constructs for the wild-type Tax protein have been described previously (7). *Escherichia coli* BL21(DE3) cells transformed with the Tax expression plasmid, pET11-TaxH6, were grown at 37 °C in 2 liters of Terrific Broth medium containing 100 $\mu\text{g}/\text{ml}$ ampicillin until $A_{600\text{ nm}} = 1.0$ –1.5. Tax expression was then induced with 40 μM isopropyl- β -D-thiogalactopyranoside at room temperature overnight. Cells were harvested and resuspended in 20 ml of 50 mM phosphate-buffered saline (pH 8.0) containing 0.3 M NaCl, 0.25 mM phenylmethylsulfonyl fluoride, 0.5 mM β -mercaptoethanol, and 10 mM imidazole. The cells were ruptured by sonication over an ice-salt bath using a Bronson sonifier mounted with a microtip. Sonication was carried out at 70% duty cycle four times at 1 min each. After centrifugation in a Sorvall SS-34 rotor at 16,000 rpm for 50 min at 4 °C, the supernatant was mixed with 5 ml of nickel-nitrilotriacetic acid-agarose (QIAGEN, Germany) at 4 °C for at least 2 h. The protein-bound gel matrix was then packed into a column (1.5 \times 10 cm), and washed with 40 ml of the same buffer containing 40 mM imidazole. Tax was then eluted with a 60 ml of gradient of 80–300 mM imidazole. Fractions containing Tax were dialyzed against buffer D (20 mM Hepes (pH 7.9), 150 mM KCl, 0.2 mM EDTA, 0.5 mM phenylmethylsulfonyl fluoride, 0.5 mM dithiothreitol, and 20% glycerol) and stored frozen at –80 °C. Protein concentrations were determined by the Bio-Rad protein assay using bovine serum albumin as a standard.

The glutathione S-transferase (GST)-bZip constructs have been described previously (8, 20). The GST-CBP_{451–682} protein containing the Tax-binding domain (residues 451–682) of human CBP, was derived

CREB-Hu	EAAAR	KREVRIMKNREAAARECRRKKK
ATF-1	DDQL	KREIRLMKNREAAARECRRKKK
CREM	EEATR	KRELRLMKNREAAARECRRKKK
CREBP1/ATF-2	EDPDE	KRRKFLERNRAAASRCRQKR
XBP1-Hu	LSPEE	KALRRKLKNRVAQAQTAARDKKAR
SKO1	EEQER	KRKEFLERNRVAASKFRKRKEYIKK
ATF-4	GEKLD	KKLKKMEQNKTAATRYRQKKR
AP1	SQERI	KAERKRMNRNIAASKCRKRKLER
JunB-Hu	DQERI	KVERKRLNRNLAATKCRKRKLER
GCN4	PESSD	PAALKRARNTEAARRSRARKLQR
Fos-Hu	PEEEE	KRRIRERNKMAAAKCRNRRRE
C/EBP	DKNSN	EYVRRRERNNIAVRKSRDKAKQR
NF-IL6 (CEBB-Hu)	DKHSD	EYKIRRRERNNIAVRKSRDKAKMR

FIG. 1. Sequence comparison of the basic domains of a select group of bZip proteins. The DNA and protein sequence of the basic domains were obtained from the National Center for Biotechnology Information (NCBI) GenBank™ sequence data base via a World Wide Web search. The sequences are arranged in accordance to their similarities to the CREB basic domain from top to bottom. The sequences are divided into three groups based on their DNA binding specificities as described in the text. The EEAAR residues of CREB bZip and corresponding aa sequences in other bZip proteins are separated from the basic domains. The NR, AA, CR residues (in *boldface*) are responsible for base-specific recognition. The residues (Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹) identified in this study to be critical for Tax binding are *underlined*. Proline residues are both *italicized* and *underlined*.

from a GST construct kindly provided by R. H. Goodman (Vollum Institute, Portland, OR). GST fusion proteins were prepared from isopropyl- β -D-thiogalactopyranoside-induced, *Escherichia coli* strain DH5 α harboring the respective expression constructs. Cells were lysed by sonication as described previously. Bacterial lysates were incubated with glutathione-Sepharose 4B (Amersham Pharmacia Biotech) at 4 °C with gentle agitation for 30 min. The protein-bound Sepharose was washed repeatedly and GST fusion proteins were eluted in reduced glutathione buffer and dialyzed overnight against buffer D at 4 °C. Proteins were analyzed by SDS-polyacrylamide gel electrophoresis (12.5% gel), and fractions were stored at –80 °C.

Gel Electrophoretic Mobility Shift Assay—A 46-bp *BglII*-*NcoI* fragment containing the promoter proximal copy of the HTLV-1 21-bp repeat was labeled with [α -³²P]dATP using Klenow enzyme and purified by electrophoresis on a 7.5% polyacrylamide gel. Protein/DNA binding reactions were carried out as previously reported (7, 8, 23, 24). The reaction mixtures (typically 8 μl) were electrophoresed on a 4.5 or 6% nondenaturing polyacrylamide gel (30:1 acrylamide to bisacrylamide, 18 \times 14 cm) in Tris-glycine-EDTA buffer (25 mM Tris, 192 mM glycine, 1 mM EDTA, pH 8.5) at 200 V and 4 °C for 2–3 h. The amounts of bZip, Tax, and GST-CBP_{451–682} proteins are approximately 40 ng, 0.1 μg and 0.1 μg , respectively, as determined by the Bradford assay or by Coomassie Blue staining followed by quantitation using an Eagle Eye II system (Stratagene Inc.). The gel was then dried on a piece of Whatman filter paper and autoradiographed. Purified Tax, bZip, and GST fusions were used in all assays with the exception that crude bacterial lysates containing bZip mutants were used in the experiments shown in Fig. 6.

Derivation of CREB Basic Domain Mutants—All mutations in the CREB basic domain were introduced by polymerase chain reactions using the Vent polymerase and specific mutagenic primers in two steps. The first PCR was carried out using the mutagenic primer and a primer for the extreme COOH-terminal coding sequence of CREB. The second PCR was carried out using, in large excess, the DNA fragment containing a given base substitution together with a DNA fragment harboring the full CREB bZip coding sequence and a suitable primer pair that allowed amplification of the aa 280–341 region of CREB. The DNA fragments containing the targeted changes were cloned into the pCRII vector, and their sequences were confirmed by sequence analyses. The mutant bZip fragments were subcloned into the pET11 vector via *NcoI* (or *NdeI*) and *BamHI* restriction endonuclease cleavage sites for expression. The sequences of the primers are as follows: for the three minimal bZip domains, CCATATGAAGAGAGAGGTCGGT (aa 285–341), CATGCCATGGCGTCCTCCCCA (aa 268–341), and ACAGCCATGGAAAGAGCAGACGA (aa 280–341) were used. For CREB bZip mutants containing the ATF-2 sequences, GCCATGGAAGAAGCAGCACGAAAAAGGAGAAAGTTTITA (EEAAR/2b/cz) and GCAGCCATGGAAATCCTGTGAAAGAGAGAGGTTTCG (EDPDE/cb/cz) were used. For the alanine substitutions and the K305L mutation, a primer pair, GCAGCCATGGAAGAAGCAGCACGA (CREB bZip upstream) and GGATCCTT-ATTA ATGGTGGTG (CREB bZip downstream), were used

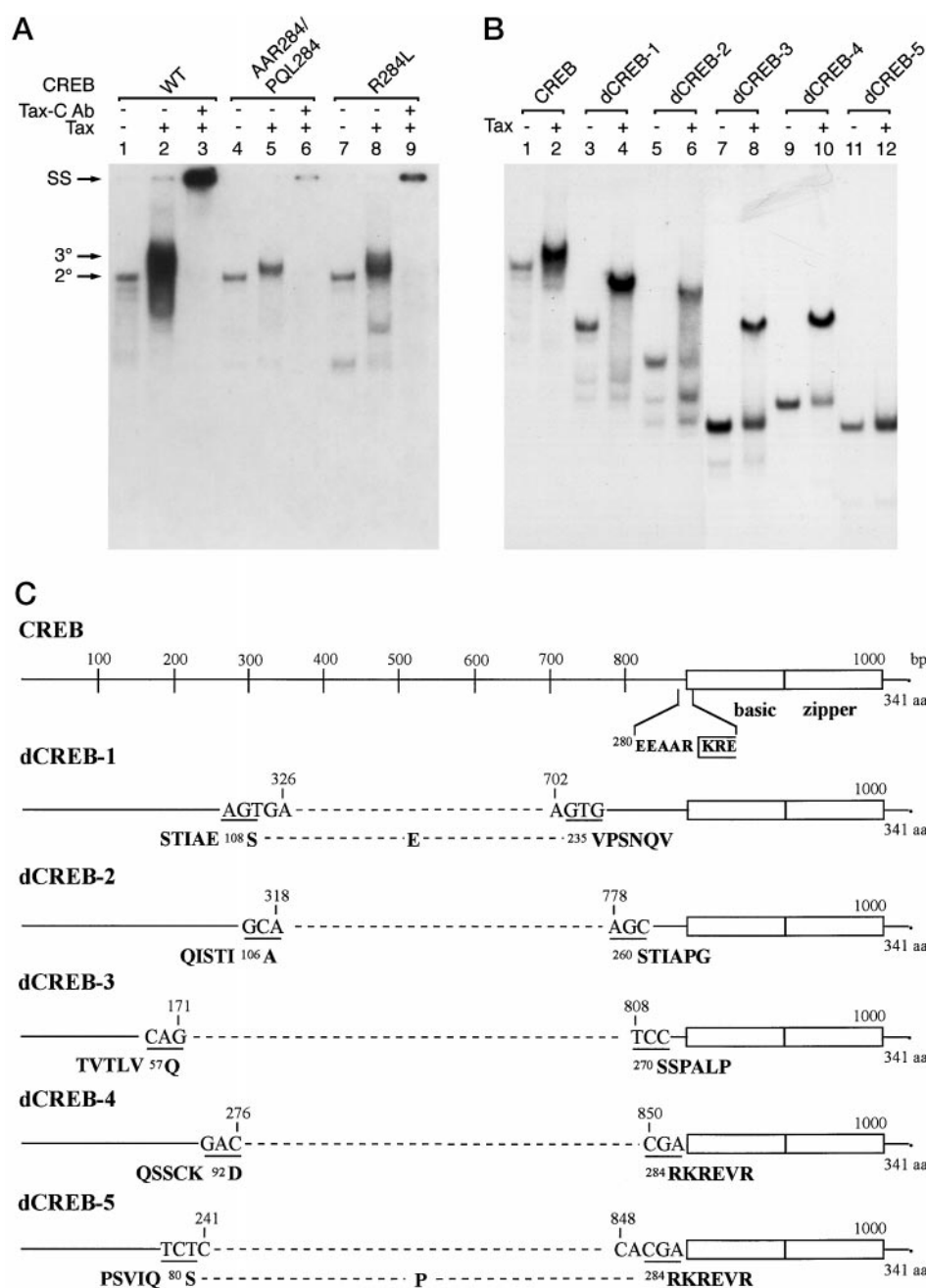


FIG. 2. Incorporation of proline residues in the vicinity of Arg²⁸⁴ of CREB disrupts Tax binding. *A*, CREB mutants, AAR284/PQL284 and R284L, with aa residues ²⁸²Ala-Ala-Arg²⁸⁴ substituted with Pro-Gln-Leu and Arg²⁸⁴ substituted with Leu, respectively, were produced by PCR, expressed, and purified from *E. coli* as described previously (8, 20). The CREB-21-bp repeat, Tax-CREB-21-bp repeat, and the TaxC-Ab supershifted complexes are denoted as 2°, 3°, and SS, respectively. The Tax binding activities of AAR284/PQL284 and R284L are 12 and 33% of wild-type as quantitated by a PhosphorImager. *B*, EMSAs for the mutants listed in *C* were carried out as reported. *C*, the CREB deletion mutants dCREB 1–5 were generated by Bal-31 deletion after cleavage of the pET11a-CREB plasmid with the *Kpn*I restriction endonuclease. The extent of the deletion and the protein sequence were deduced from DNA sequence analyses. For each deletion, the nucleotide position numbers of the CREB cDNA (*top*) and the amino acid sequence across the deletion junction (*bottom*) are indicated.

in conjunction with the mutagenesis primers to produce the coding sequences for the mutant bZips (aa 280–341). The mutagenesis primers were GCAGCACGAGCAAGAGAGGTC (K285A), CACGAAAGCGGAG-GTTCG (R286A), AAGAGAGCGGTCCGT (E287A), AAGAGAGAGG-CCGTCTAATG (V288A), AGAGAGGTTGCTCTAATGAAG (R289A), GAGGTTCTGTCGATGAAGAAC (L290A), AGGTCCGTCTAGCGAA-GAACAG (M291A), TCCGTCTAATGGCGAACAGGGAAGC (K292A), AACAGGGCAGCAGCT (E295A), GAAGCAGCTGCTGAGTGTCTGTA-GAA (R298A), GCTCGTGCTGTCGT (E299A), GAGTGTCTGCAAA-GAAGAAAGAA (R302A), GTGTCTGAGAGCGAAGAAAGAA (K303A), CGTAGAAAGCGAAAGA ATATG (K304A), and CGTAGAAA-GAAGCTAGAATATGTG (K305L).

Molecular Modeling of the CREB-HTLV Enhancer Complex—The CREB-HTLV enhancer model is based on the x-ray crystal structure of

the GCN4 bZIP domain bound to a DNA containing a CRE site (5'-TGACGTCA-3') (34, 36). An additional α -helical turn was modeled at the N terminus of the GCN4 basic region to accommodate Arg²⁸⁴, which is required for the interaction of CREB with Tax. The model was manipulated with the program O previously reported (37) and Fig. 7 was created using RIBBONS (38).

RESULTS AND DISCUSSION

Disruption of an Extended α -Helical Structure Formed by the CREB Basic Domain and Amino Acid Residues to Its Immediate NH₂ Terminus Abrogates Tax Binding—X-ray structures of the bZip domains of GCN4 and c-Jun/c-Fos (33–36) show the leucine zipper and the basic domain to form a continuous

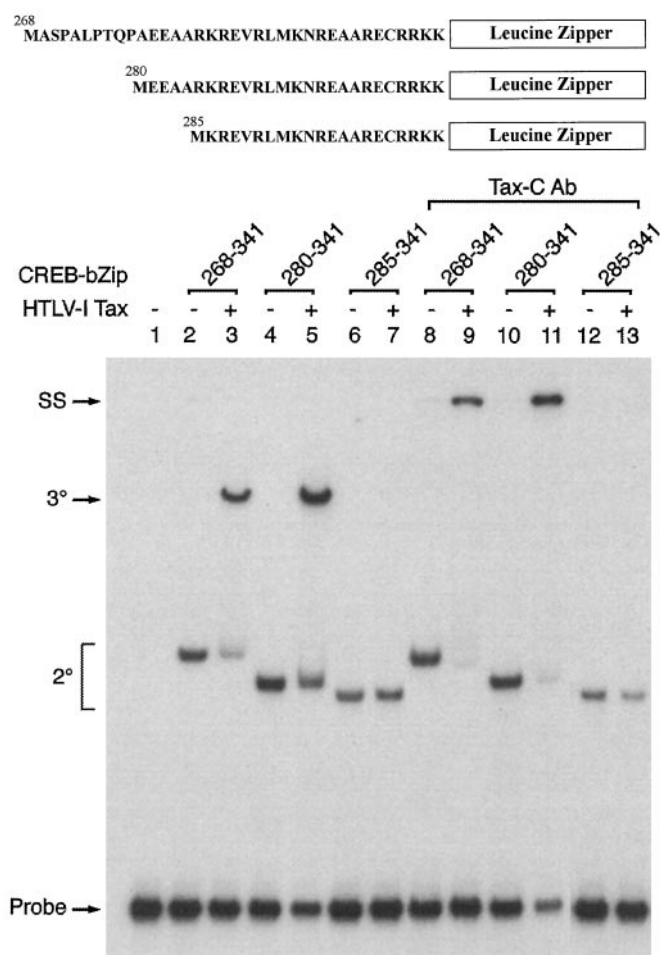


FIG. 3. Minimal CREB bZip domain for Tax binding consists of aa 280–341. Amino acid positions of the basic regions of three CREB bZip derivatives are shown (top). The promoter proximal HTLV-I 21-bp repeat was ^{32}P -labeled, gel-purified, and used as a probe.

α -helix. The immediate NH_2 termini of the basic domains of GCN4 and c-Fos contain proline residues and were not included in the resolved bZip structures (Fig. 1). Other bZip members such as CREB, ATF-1, ATF-2, and XBP-1 also contain proline residues at variable distances in the immediate NH_2 termini of their respective basic domains (Fig. 1). These proline residues are expected to disrupt the continuity of the α -helical structure.

As reported previously, ATF-1, which contains DDPQL residues (instead of EEAAR found in CREB) to the NH_2 terminus of its basic domain, interacts relatively weakly with Tax (20). This reduced affinity appears to be due to a combination of both the proximity of the proline residue to the basic domain and the substitution of a Leu for the Arg²⁸⁴ residue in CREB. This conclusion is supported by results from electrophoretic mobility shift assays (EMSA) showing that the replacement of Arg²⁸⁴ with Leu reduced the Tax binding activity to 30% that of the wild-type CREB (compare Fig. 2A, lanes 1–3 with lanes 7–9, also see Adya *et al.* (20)) while substitution of ²⁸²AAR²⁸⁴ residues with PQL further reduced Tax binding to 10% of the wild-type level (Fig. 2A, lanes 4–6) as judged by the amounts of the ternary complex (lanes 2, 5, and 8) and the Tax-COOH-terminal antibody (Tax-C Ab) supershift complex (lanes 3, 6, and 9). These results had previously been confirmed by co-immunoprecipitation (20). In the course of defining the minimal CREB domain required for Tax binding, a series of CREB mutants with internal deletions have been generated (Fig. 2C). As expected, deletions to the NH_2 terminus of the EEAAR

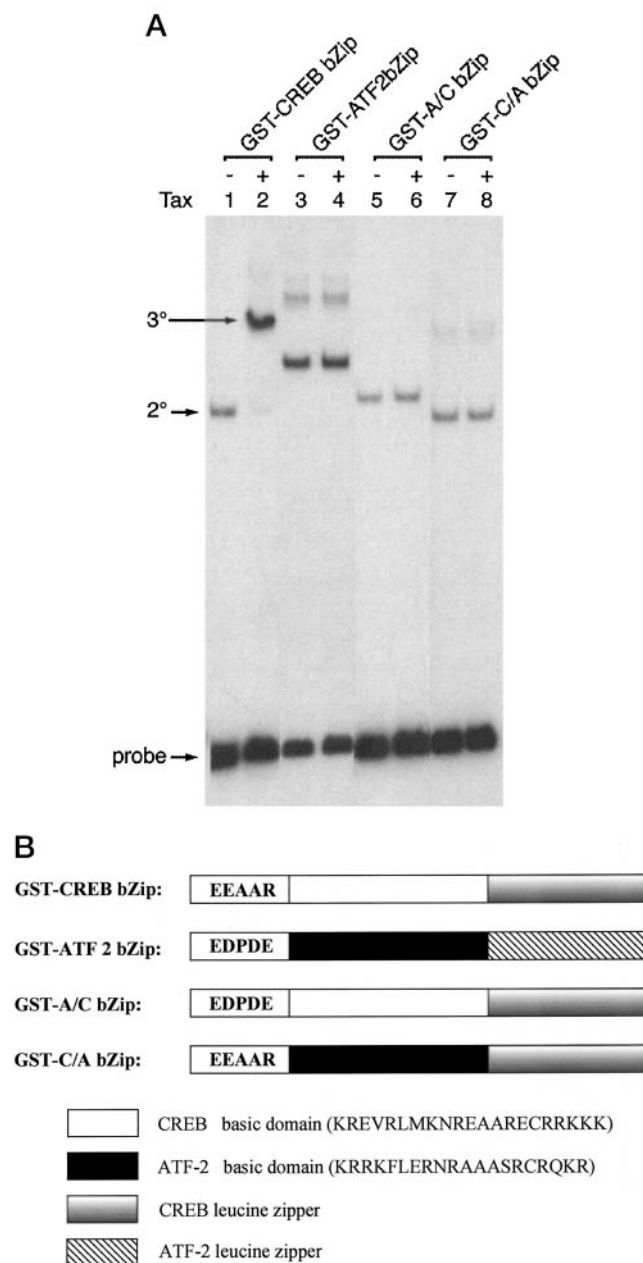


FIG. 4. EEAAR residues in CREB bZip are necessary but not sufficient for Tax binding. A, the domain switching between CREB and ATF-2 basic regions was carried out by hybrid extension PCR and confirmed by DNA sequence analyses. GST-A/C bZip contains EDPDE in place of EEAAR in the backbone of the CREB bZip, while GST-C/A bZip contains EEAAR in place of EDPDE in the backbone of the ATF-2 bZip. B, a schematic diagram of the recombinant GST-bZip proteins analyzed in A.

residues (Fig. 2, B and C, dCREB 1–3) covering the region spanning the kinase phosphorylation site, Ser¹³³, had no effects on Tax binding (Fig. 2B, lanes 1–8). Two of the mutants, dCREB-4 and dCREB-5, are particularly noteworthy (Fig. 2B, lanes 9–12). In both mutants, the aa sequence that immediately abuts the NH_2 terminus of Arg²⁸⁴ was significantly altered. In dCREB-4, aa residues (SSCKD) became incorporated next to Arg²⁸⁴ (Fig. 2C). These residues were expected to maintain an α -helical conformation based on secondary structure predictions and did not affect the interaction of the basic domain with Tax (Fig. 2B, lanes 9 and 10). In contrast, in dCREB-5, a proline residue was introduced immediately adjacent to Arg²⁸⁴ (Fig. 2C). This did not affect CREB bZip binding

to the 21-bp repeat DNA probe but completely abrogated Tax binding as evidenced by the absence of the Tax-CREB-probe ternary complex in the EMSA (Fig. 2*B*, lanes 11 and 12). From these results, we conclude that an uninterrupted α -helical structure covering aa residues ²⁸⁰EEAAR²⁸⁴ is an important determinant for Tax binding.

Defining the Minimal CREB bZip Domain Required for Tax Binding—To delimit the minimal CREB bZip structure required for Tax binding, three derivatives of CREB bZip containing aa residues 268–341, 280–341, and 285–341 of CREB, respectively, were generated. All three bZip domains bound the DNA probe containing the promoter proximal HTLV-I 21-bp repeat with similar affinities. CREB-bZip 268–341 and 280–341 were able to interact with Tax as evidenced by ternary complex formation (Fig. 3, lanes 3 and 5). The presence of Tax in the ternary complex was demonstrated by a supershift with an antibody, Tax-C Ab, directed against the COOH-terminal 33 aa of Tax (Fig. 3, lanes 9 and 11). The absence of the ²⁸⁰EEAAR²⁸⁴ residues, while having little effect on DNA binding (Fig. 3, lane 6), rendered CREB bZip 285–341 defective in its interaction with Tax (Fig. 3, lanes 7 and 13). These results are also consistent with the data in Fig. 2*B* where protein sequences to the NH₂ terminus of the EEAAR residues were found to be unimportant for Tax binding. These results indicate that CREB bZip aa 280–341 constitutes the minimal domain for ternary complex formation.

Consistent with the results in Fig. 3, a replacement of ²⁸⁰EEAAR²⁸⁴ with the EDPDE sequence of the ATF-2 bZip, a protein that does not bind Tax (Fig. 4, lanes 3 and 4), abrogated Tax binding (Fig. 4, lanes 5 and 6). These data support the conclusion that the ²⁸⁰EEAAR²⁸⁴ residues immediately upstream of the CREB basic domain are required for strong Tax binding. The EEAAR residues alone are not sufficient, however. Other parts of the basic domains of CREB and ATF-1 also contribute to the interaction with Tax. Indeed, a recombinant ATF-2 bZip containing the EEAAR residues in place of the corresponding aa residues (EDPDE) in the native ATF-2 bZip also lacked the Tax binding activity (Fig. 4, lanes 7 and 8). This is in agreement with earlier data showing that, although ATF-1 lacks the ²⁸⁰EEAAR²⁸⁴ residues (see Fig. 1 for the aa sequence of the ATF-1 basic domain), it nonetheless binds Tax with an affinity approximately one-tenth that of CREB (20).

The transcriptional co-activators, CBP/p300, have recently been shown to interact with Tax (31, 32). Analyses of various GST-CBP fusions have identified aa 451–682 of CBP to contain the domain important for binding Tax. Indeed, although a GST-CBP_{451–682} fusion did not interact directly with CREB bZip (Fig. 5, lanes 5–7), it stabilized the interaction between Tax, CREB bZip, and the HTLV-I 21-bp repeat to produce a quaternary complex that could readily be detected by EMSA (Fig. 5, lanes 8 and 9). The ability of CREB bZip mutants to support quaternary complex formation correlated directly with their ability to form ternary complexes. This is as indicated by EMSA showing that CREB-bZip 268–341 and 280–341 but not CREB-bZip 285–341 supported the formation of the quaternary complex (Fig. 5, lanes 8–10). Although it is possible to observe stable ternary complexes with the CREB-bZip 268–341 and 280–341 proteins under optimal conditions (Fig. 3), the ternary complex that Tax formed with CREB-bZip is not as stable as that formed with the full-length CREB or GST-CREB-bZip under EMSA conditions. Inclusion of GST-CBP_{451–682}, however, strengthens the Tax/CREB-bZip interaction to yield readily detectable quaternary complexes (Fig. 5, lanes 8 and 9). For this reason, subsequent analyses of Tax/CREB-bZip interactions were carried out in the presence of GST-CBP_{451–682}.

Specific aa Residues in the CREB Basic Domain Are Impor-

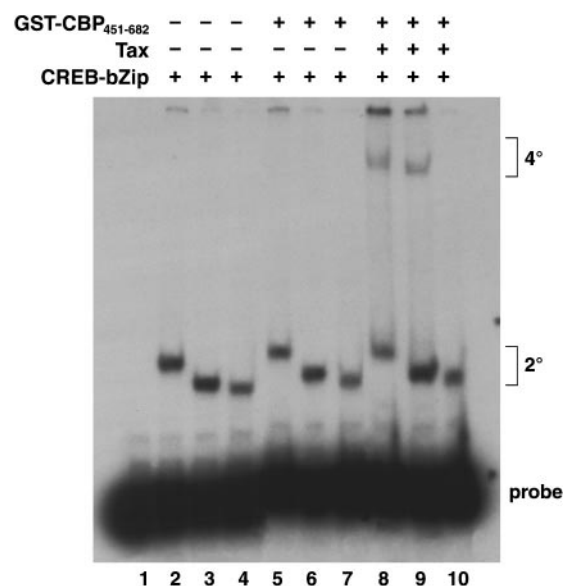


FIG. 5. Quaternary CBP-Tax-CREB-bZip-21-bp repeat complex formation correlates with Tax binding. EMSAs were carried out as in Figs. 2 and 3, except that a GST-CBP_{451–682} fusion was included in reactions 5–10 and Tax was added in lanes 8–10. CREB-bZip aa 268–341, aa 280–341, and aa 285–341 (as in Fig. 3) were used in each set of three in lanes 2–4, 5–7, and 8–10.

tant for Tax Binding—To locate the aa residues in the CREB basic domain that are important for Tax binding, we began by targeting alanine substitutions to three glutamate residues, Glu²⁸⁷, Glu²⁹⁵, and Glu²⁹⁹. These residues were chosen because of the long side chain and the negative charge of glutamate, which we reasoned might be important for protein-protein contacts. The alanine substitutions were introduced into the CREB-bZip 280–341 backbone (Fig. 6*A*). These mutations did not affect CREB-bZip binding to DNA overtly except for E287A, which reduced the bZip DNA binding activity to approximately 1/2 that of the wild-type (Fig. 6*B*, lanes 5–7). E287A also altered the mobility of the bZip domain both in SDS-polyacrylamide gel electrophoresis and in EMSA (Fig. 6, *A* and *B*, lane 5). While E287A and E295A had little effect on bZip-Tax binding as judged by the quaternary complexes (Fig. 6*C*, lanes 3–5), E299A drastically inactivated this interaction (Fig. 6*C*, lane 5). A helical wheel representation of aa 282–304 of the CREB basic domain reveals both Arg²⁸⁴ and Glu²⁹⁹ to be located on the same surface of the α -helix, opposite to the highly conserved residues, Asn²⁹³, Arg²⁹⁴, Ala²⁹⁶, Ala²⁹⁷, and Cys³⁰⁰ that are involved in base-specific DNA recognition (Fig. 6*D*).

To exhaustively identify aa residues in the CREB basic domain that are engaged in Tax binding, additional alanine substitutions were carried out for residues Arg²⁸⁶, Arg²⁸⁹, Leu²⁹⁰, Lys³⁰³, Lys³⁰⁴, and Lys²⁸⁵, Val²⁸⁸, Met²⁹¹, Lys²⁹², Arg²⁹⁸, Arg³⁰² located on the same, and opposing side, of the DNA binding surface, respectively (Fig. 6*D*). Because the invariant aa residues are not responsible for Tax binding as evidenced by the lack of interaction between Tax and ATF-2, no attempt was made to alter these residues. We took advantage of the fact that the GST-CBP_{451–682} fusion stabilized the multiprotein complex on the 21-bp repeat and used the quaternary complex formation as a measure of CREB bZip interaction with Tax. Consistent with the notion that the highly conserved residues are principally responsible for DNA binding, none of the alanine mutants showed significant alterations in their affinities for the 21-bp repeat-containing DNA probe (Fig. 6*B*, lanes 4–19). Further, most aa substitutions did not significantly affect Tax binding, except for Glu²⁹⁹ and Met²⁹¹, whose substitutions with alanine residues abrogate Tax binding and quaternary complex

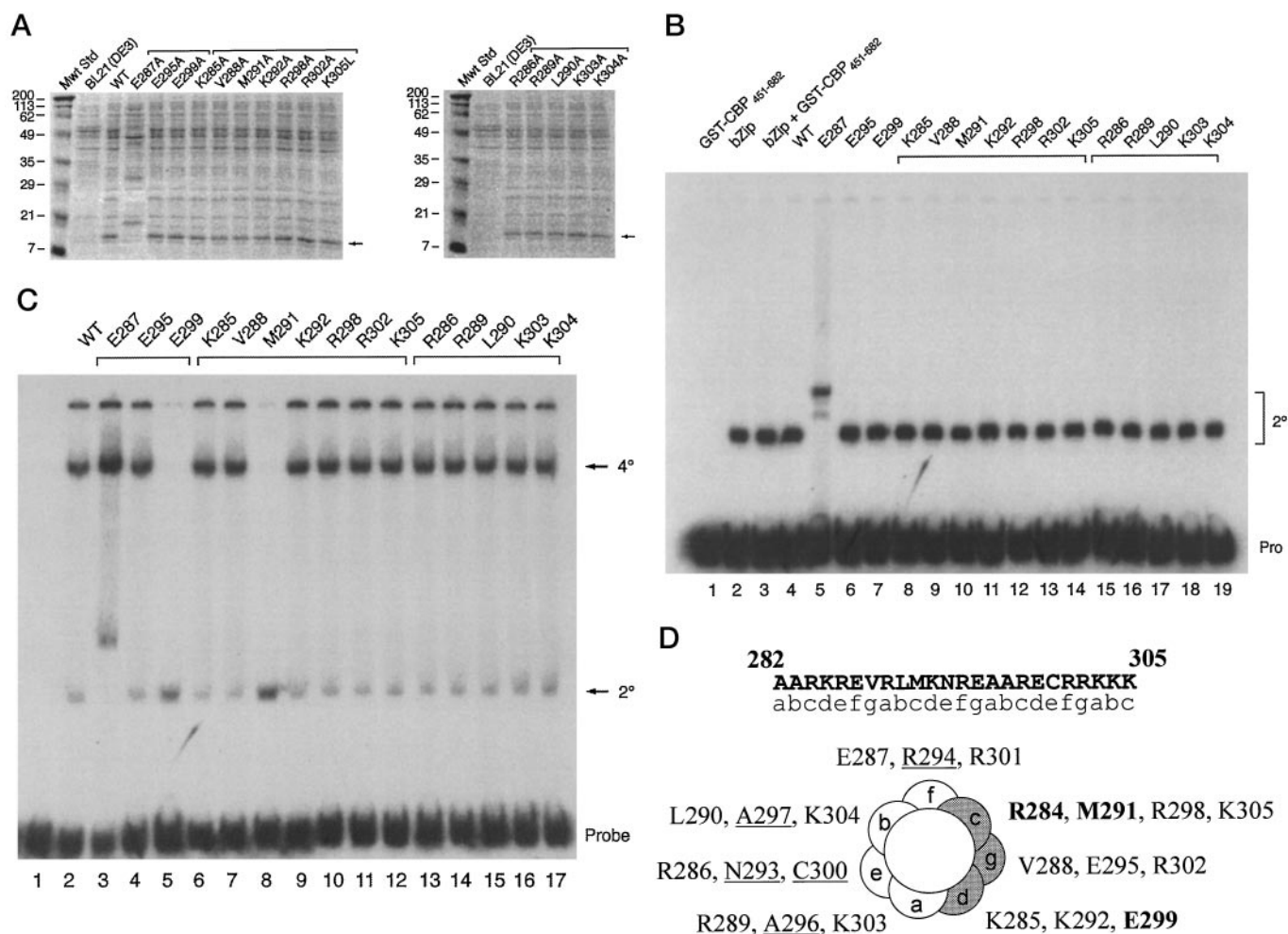


FIG. 6. Identification of aa residues in the CREB basic domain that are critical for Tax binding. CREB bZip mutants were derived by PCR and constructed in the background of the minimal CREB aa 280–341 bZip cloned in the pET11d vector. Bacterial lysates expressing comparable amounts of the respective mutant bZip proteins (panel A, marked by an arrow, except for E287A, which migrated anomalously slower) were used together with the p^{32} -labeled 21-bp repeat DNA probe in EMSA either alone (B), or with the purified Tax and the GST-CBP_{451–682} fusion (C). D, helical wheel representation of the CREB basic domain and its neighboring region. The conserved amino acid residues important for DNA contacts (underlined) are all located on one side (unfilled circles), while the Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹ (boldface) are positioned on the opposing surface of the α -helix (shaded circles).

formation (Fig. 6C, lanes 5 and 8).

An examination of the aa sequence of a select group of bZip domains reveals that Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹ are unique to CREB (Fig. 1). Another member of the CREB subfamily of bZip proteins, CREM, contains identical aa residues at the corresponding positions, while ATF-1 contains a leucine at the residue corresponding to Arg²⁸⁴. All of these proteins have been shown to interact with Tax (8, 9, 11, 12, 20).

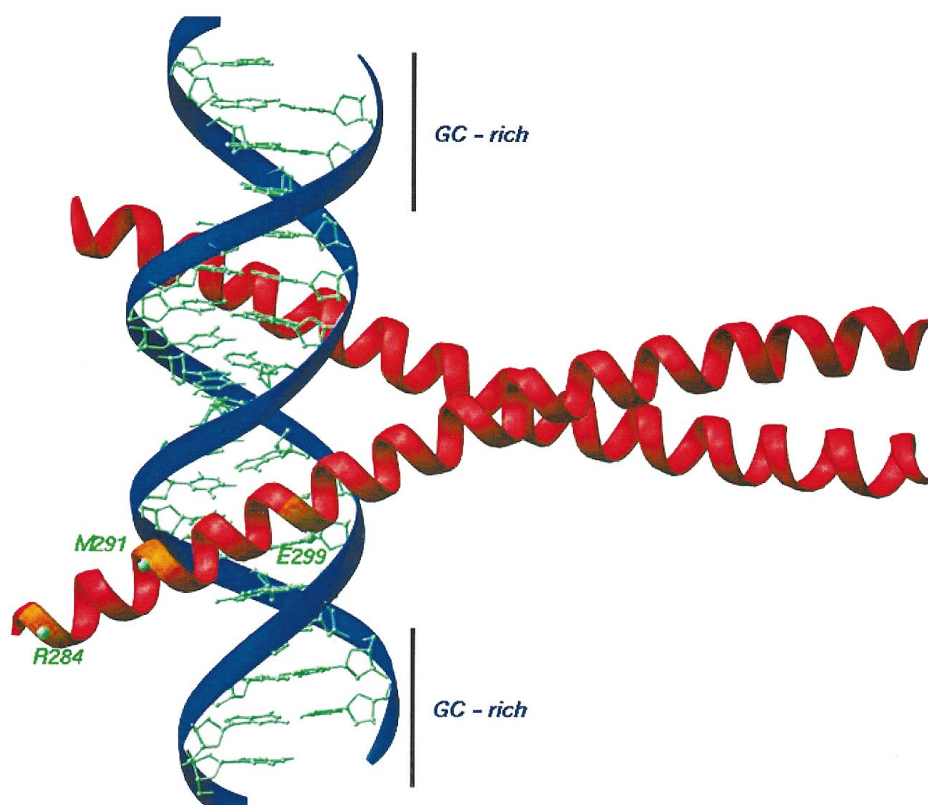
Molecular Modeling of CREB bZip Based on the Structure of GCN4 bZip-CRE Complex—We have used the GCN4-CRE crystal structure to model CREB bound to an HTLV 21-bp repeat to understand how CREB and Tax interact when bound to DNA (34, 36). Three critical assumptions are made in our modeling. First, we assume that differences in protein and DNA sequence between the CREB and GCN4 complexes do not significantly affect the alignment of the CREB basic region in the major groove of the DNA. Indeed, all residues that contact base pairs in the major groove of the DNA are conserved between GCN4 and CREB, as are many of the positively charged amino acid side chains that help anchor the basic region in the major groove through electrostatic interactions with the phosphate backbone. We also assume that interactions with Tax will not perturb the way in which the CREB basic region interacts with the DNA. Finally, we have modeled the CREB basic region to

have at least one additional turn of α -helix at its NH₂ terminus compared with GCN4 based on the secondary structure preferences of the residues in this region.

The mutagenesis data indicate that Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹ make critical contacts with Tax. Met²⁹¹ and Glu²⁹⁹ are positioned on the same side of the basic region in a helix extending away from the DNA and conceivably accessible to Tax (Fig. 7). Arg²⁸⁴ lies just two residues beyond the NH₂ terminus of the GCN4 basic region α -helix. In GCN4, this helix is broken at its NH₂ terminus by a proline residue (Fig. 1). Residues in this region of CREB favor a helical conformation. Further, as shown in Figs. 1 and 2, the introduction of proline residues nearby severely impaired the binding activity for Tax. Thus, Arg²⁸⁴ is most likely in a helical conformation situated on the same side of the basic region helix as Met²⁹¹ and Glu²⁹⁹, defining an approximately 23 Å (1.5 Å/residue for α -helix) long Tax contact surface (Fig. 7).

It is worth noting that the amino acid residue corresponding to Arg²⁸⁴ of the CREB basic domain is dispensable for DNA binding (Fig. 2). This region was not present in the GCN4 bZip crystal structure. The modeling of Arg²⁸⁴ is based on the assumption that Arg²⁸⁴ and its neighboring region are parts of the continuous α -helix formed by the bZip domain (in red). Arg²⁸⁴, Met²⁹¹, and Glu²⁹⁹, residues that are critical for Tax

FIG. 7. A model of a CREB homodimer bound to DNA illustrating amino acid residues that contact Tax. The CREB-CRE complex is based on the x-ray crystal structure of the GCN4 bZIP domain bound to a CRE site (36). The CREB main chain is displayed as a red ribbon with green highlights and spheres at the positions of residues in one of the CREB basic regions that contact Tax. GC-rich sequences that flank the CRE motif and facilitate binding of Tax are also indicated.



binding, are located on the same surface of the helix and extend outward from the major groove for protein contacts. The CRE DNA (in blue) takes on characteristics of the A-DNA and has a deeper major groove, as previously observed for the GCN4 bZip-CRE complex.

How the 5' G-rich/3' C-rich Sequences Flanking the CRE Might Effect a Specific Interaction between CREB and Tax—*In vitro* selection of DNA elements that bind Tax/CREB with high affinity conclusively demonstrates that runs of guanosine 5' or cytosine 3' to the CRE, are required to recruit Tax to a CREB-DNA complex. The minor groove of the poly(G)-poly(C) flanking sequence (Fig. 7, the region marked as *G/C-rich*) is adjacent to the Tax contact surface on the CREB basic domain in our model. It is likely, therefore, that Tax bound to the outer edge of the CREB basic region helix could also contact the minor groove of the G/C-rich sequences. The earlier finding that Tax/CREB and CREB alone have identical DNase I and dimethyl sulfate footprints was interpreted to suggest that Tax makes little direct contact with the DNA in its major groove (15). However, these reagents do not reveal contacts between Tax and the G/C-rich minor groove of the 21-bp repeat, as DNase I tends not to cut across GC-rich minor grooves and dimethyl sulfate alkylates only G residues on the N7 atom exposed in the major groove (39). Using the minor groove-specific, footprinting reagent, Fe-methidiumpropyl-EDTA, Lenzmeier *et al.* (19) have recently shown that Tax expands the footprint made by CREB. This suggests that, upon binding to the CREB basic domain, Tax indeed contacts the minor groove of the GC-rich regions. Chemical cross-linking results also support this to be the case (19). In addition to a direct contact with Tax, we think the G/C-rich sequences may also alter the DNA structure of the CRE, which, in turn, may modulate the conformation of the CREB basic domain to produce a structure suitable for Tax binding. Because the contact surface for Tax spans almost the entire length of the basic domain helix, alterations of the helical structure as a function of the sequences flanking the CRE motif may determine if stable interaction

with Tax can occur; thereby contributing to the DNA specificity of the ternary complex assembly. A definitive understanding of how the Tax-CREB complexes specifically recognize the HTLV 21-bp repeat elements must await structural analysis of this system.

Acknowledgment—We thank Dr. R. Goodman for the gift of the GST-CBP_{451–682} construct.

REFERENCES

1. Cann A. J., and Irvin S. Y. Chen (1996) *Virology* (Bernard N. Fields, ed) pp. 1849–1880, Lippincott-Raven, Philadelphia.
2. Paskalis, H., Felber, B. K., and Pavlakis, G. N. (1986) *Proc. Natl. Acad. Sci. U. S. A.* **83**, 6558–6562.
3. Rosen, C. A., Park, R., Sodroski, J. G., and Haseltine, W. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84**, 4919–4923.
4. Brady, J., Jeang, K. T., Duvall, J., and Khoury, G. (1987) *J. Virol.* **61**, 2175–2181.
5. Fujisawa, J., Toita, M., and Yoshida, M. (1989) *J. Virol.* **63**, 3234–3239.
6. Giam, C. Z., and Xu, Y. L. (1989) *J. Biol. Chem.* **264**, 15236–15241.
7. Zhao, L. J., and Giam, C. Z. (1991) *Proc. Natl. Acad. Sci. U. S. A.* **88**, 11445–11449.
8. Zhao, L. J., and Giam, C. Z. (1992) *Proc. Natl. Acad. Sci. U. S. A.* **89**, 7070–7074.
9. Suzuki, T., Fujisawa, J. I., Toita, M., and Yoshida, M. (1993) *Proc. Natl. Acad. Sci. U. S. A.* **90**, 610–614.
10. Goren, I., Semmes, O. J., Jeang, K. T., and Moelling, K. (1995) *J. Virol.* **69**, 5806–5811.
11. Bantignies, F., Rousset, R., Desbois, C., and Jalinot, P. (1996) *Mol. Cell. Biol.* **16**, 2174–2182.
12. Shnyreva, M., and Munder, T. (1996) *J. Virol.* **70**, 7478–7484.
13. Derse, D. (1987) *J. Virol.* **61**, 2462–2471.
14. Montagne, J., Beraud, C., Crenon, I., Lombard Platet, G., Gazzolo, L., Sergeant, A., and Jalinot, P. (1990) *EMBO J.* **9**, 957–964.
15. Paca Uccaralertkun, S., Zhao, L. J., Adya, N., Cross, J. V., Cullen, B. R., Boros, I. M., and Giam, C. Z. (1994) *Mol. Cell. Biol.* **14**, 456–462.
16. Yin, M. J., and Gaynor, R. B. (1996) *Mol. Cell. Biol.* **16**, 3156–3168.
17. Yin, M. J., Paulsen, E., Seeler, J., and Gaynor, R. B. (1995) *J. Virol.* **69**, 6209–6218.
18. Boros, I. M., Tie, F., and Giam, C. Z. (1995) *Virology* **214**, 207–214.
19. Lenzmeier, B. A., Giebler, H. A., and Nyborg, J. K. (1998) *Mol. Cell. Biol.* **18**, 721–731.
20. Adya, N., Zhao, L. J., Huang, W., Boros, I., and Giam, C. Z. (1994) *Proc. Natl. Acad. Sci. U. S. A.* **91**, 5642–5646.
21. Yin, M. J., and Gaynor, R. B. (1996) *J. Mol. Biol.* **264**, 20–31.
22. Perini, G., Wagner, S., and Green, M. R. (1995) *Nature* **376**, 602–605.
23. Tie, F., Adya, N., Greene, W. C., and Giam, C. Z. (1996) *J. Virol.* **70**, 8368–8374.
24. Adya, N., and Giam, C. Z. (1995) *J. Virol.* **69**, 1834–1841.
25. Hai, T. W., Liu, F., Coukos, W. J., and Green, M. R. (1989) *Genes Dev.* **3**,

- 2083–2090
26. Gonzalez, G. A., and Montminy, M. R. (1989) *Cell* **59**, 675–680
27. Gonzalez, G. A., Yamamoto, K. K., Fischer, W. H., Karr, D., Menzel, P., Biggs, W., Vale, W. W., and Montminy, M. R. (1989) *Nature* **337**, 749–752
28. Sun, P., Lou, L., and Maurer, R. A. (1996) *J. Biol. Chem.* **271**, 3066–3073
29. Chrivia, J. C., Kwok, R. P., Lamb, N., Hagiwara, M., Montminy, M. R., and Goodman, R. H. (1993) *Nature* **365**, 855–859
30. Kwok, R. P., Lundblad, J. R., Chrivia, J. C., Richards, J. P., Bachinger, H. P., Brennan, R. G., Roberts, S. G., Green, M. R., and Goodman, R. H. (1994) *Nature* **370**, 223–226
31. Kwok, R. P., Lurance, M. E., Lundblad, J. R., Goldman, P. S., Shih, H., Connor, L. M., Marriott, S. J., and Goodman, R. H. (1996) *Nature* **380**, 642–646
32. Giebler, H. A., Loring, J. E., van Orden, K., Colgin, M. A., Garrus, J. E., Escudero, K. W., Brauweiler, A., and Nyborg, J. K. (2) (1997) *Mol. Cell. Biol.* **17**, 5156–5164
33. Ellenberger, T. E., Brandl, C. J., Struhl, K., and Harrison, S. C. (1992) *Cell* **71**, 1223–1237
34. Konig, P., and Richmond, T. J. (1993) *J. Mol. Biol.* **233**, 139–154
35. Glover, J. N., and Harrison, S. C. (1995) *Nature* **373**, 257–261
36. Keller, W., Konig, P., and Richmond, T. J. (1995) *J. Mol. Biol.* **254**, 657–667
37. Jones, T. A., Zou, J. Y., Cowan, S. W., and Kjeldgaard (1991) *Acta Crystallogr. Sec. A* **47**, 110–119
38. Carson, M. (1987) *J. Mol. Graphics* **5**, 103–106
39. Ogata, R. T., & Gilbert, W., 2. (1979) *J. Mol. Biol.* **132**, 709–728